Characterization of electroencephalographic signals using discrete wavelet transform as a tool to support the diagnosis of attention deficit hyperactivity disorder ADHD
Caracterización de señales electroencefalográficas utilizando la transformada wavelet discreta como herramienta para apoyar el diagnóstico del trastorno por déficit de atención e hiperactividad TDAH

A Thesis Presented for the Electrical Engineering Degree

Student: Julián David Pastrana Cortés
Director: Álvaro Ángel Orozco Gutiérrez, PhD
Co-director: David Augusto Cárdenas Peña, PhD

Universidad Tecnológica de Pereira
Engineering Faculty - Electrical Engineering Program
Research Group in Automática

Pereira, Risaralda, Colombia
2021

Abstract

Despite arising in childhood, attention deficit hyperactivity disorder (ADHD) can persist into adulthood, compromising the individual's social skills. ADHD diagnosis is a real challenge due to its dependence on the clinical observation of the patient, the information provided by parents and teachers, and the clinicians' expertise. Therefore, there is great interest in studying objective biomarkers extracted from electroencephalographic (EEG) signals supporting accurate diagnoses. However, the non-stationarity and non-linearity characteristics of the EEG hinders the development of such tools. This document presents a methodology for supporting the ADHD diagnosis by extracting features from the Discrete Wavelet transform of EEG signals. Due to the failed inhibitory control symptom, was consider EEG signals recorded under the Reward Stop Signal Task paradigm. Then, the logistic regressor classifier learns the linear boundary discriminating between ADHD and healthy control subjects. As a benefit, the weighting vector supporting the classification also provides a straightforward interpretation of features in spatial locations and decomposition levels. The cross-validated classification results prove that the approach reaches an F1 score of 96% on a dataset of 64 children. Besides, the interpretability results support the hypothesis that the motivational effect leads the poor impulse control in ADHD.

Aknowledgments

To God and automatic group of UTP.

Contents

List of Figures

List of Tables

Chapter 1

List of Symbols and Abbreviations

1.1 Symbols

Symbol	Definition	Symbol	Definition
t	Scalar and real variable	$\langle x(t), y(t) \rangle$	Dot product between x and y functions
$f(t)$	Function with variable t	$a_{j,k}$	Approximations coefficients at j scale
$\delta[k]$	Discrete Dirac delta function	$a_{j,k}$	Detail coefficients at j scale
$\phi(t)$	Scaling function	$Span\{f_n(t)\}$	Space spanned by $f_n(t)$ set
$\psi(t)$	Wavelet function	$\boldsymbol{X}$	Matrix
L^2	Square-integrable functions space	$\boldsymbol{x}$	Vector
$\mathbf{0}$	Zero function	$\boldsymbol{x}^\top$	Transpose vector

1.2 Abbrevations

Abb	Definition	Abb	Definition
FFT	Fast Fourier Transform	SSRT	Stop-signal response time
DFT	Discrete Fourier Transform	DC	Decreasing Condition
STFT	Short Time Fourier Transform	IC	Increasing Condition
DWT	Discrete Wavelet Transform	DB4	Daubechie 4 Wavelet
ADHD	Attention Deficit Hyperactivity Disorder	EEG	Electroencephalography

Chapter 2

Introduction

2.1 Problem statement

Attention deficit hyperactivity disorder (ADHD) is the most common chronic disease in childhood with a high probability of persistence throughout the lifespan. ADHD can lead to learning difficulties, low self-esteem, substance abuse, and delinquent activities in adolescents. Currently, ADHD is diagnosed by collecting information from children, parents and teachers, identifying clinical symptom criteria described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). However, this analysis incurs high rates of overdiagnosis due to the overlap of attentional and behavioral symptoms with other disorders.[1] Therefore, strategies have been proposed that, in conjunction with the regular clinical evaluation of ADHD, can help the specialist in a specific and sensitive diagnosis of the pathology.[2]

There is great motivation to identify low-cost and easily accessible biomarkers associated with ADHD. Thus, electroencephalography (EEG) recordings are ideal because they allow to better understand neuropathologies through physiological brain responses.[3] A typical approach is to decompose the EEG signals into frequency bands such as the theta-beta power ratio (TBR) that contrasts slow waves (4-7 Hz) and fast waves (13-30 Hz), under the assumption of more powerful slow waves in ADHD than in resting-state controls. Also, other studies have found a significant decrease in alpha band energy in the ADHD group compared to controls, especially in those electrodes located in the left frontal region, along with changes in the theta band that may differentiate ADHD patients in terms of gender.[4,5] Therefore, changes in the energy level of EEG bands, within the performance of a cognitive task, allow identifying characteristics associated with the disorder.

Although several studies focus on EEG signals frequency analysis as the Fourier transform, there are difficulties due to non-stationarity and non-linearity of the signal features.[6–8] These drawbacks can be solved through the short-time Fourier transform (STFT) but the window length can lead to poor resolutions. In addition, these signals have several frequency components including noise, so is necessary to use analysis methodologies in the

time and frequency domain.[9,10] This limitation is overcome by introducing the discrete wavelet transform (DWT), which decomposes the signal into other signals of different scales with different time and frequency resolutions.[11] Besides, DWT can be adapted to the frequency content of the examined patterns, thus leading to an optimal time-frequency resolution across wide frequency ranges allowing a better understanding of electrophysiological changes.[12,13]
In this context, this document presents a tool to support ADHD diagnosis using features extracted from the DWT decomposition of the EEG signals. The methodology was evaluated using EEG recordings under the influence of the Reward Stop Signal Task (RSST), which takes into account the executive inhibition modulation through motivation, as behavioral and electrophysiological landmarks associated with the disorder. For classification purposes, was applied a linear model known as logistic regression classifier that facilitates the interpretability of the results from the weights obtained for each feature, which allows predicting the class from its conditional probability.[14]
Thus, a multiresolution analysis of the EEG signals through the use of logistic regression classifier makes it possible to identify relevant characteristics associated with ADHD, which in future studies will facilitate a sensitive detection to the pathology, avoiding the degradation of patients. In this way, the following research question is established: What are the multiresolution characteristics of brain electrical activity that are associated with ADHD after feed a logistic regression classifier?

2.2 Justification

ADHD is a childhood neurodevelopmental disorder characterized by persistent patterns of generalized inattention, impulsivity, and/or hyperactivity that frequently interfere with normal child development.[15] The diagnosis of this pathology is made more easily during the school stage, when the symptoms may become more noticeable due to the repercussions on the child's academic activities. Symptoms can persist into adulthood and can significantly alter the individual's work and social life. Likewise, the disorder can be associated with learning difficulties, low self-esteem, substance abuse in adolescent patients, the propensity to accidents, social integration problems, and even delinquent behaviors, among other comorbidities.[16] Consequently, ADHD has a high social impact, is closely associated with impulsivity, and has a high prevalence in Colombia, particularly in Antioquia and the Eje Cafetero (15% to 17%). It is also important to note that of the three subtypes of ADHD (predominantly impulsive-hyperactive, inattentive, or combined) in the Colombian population, the combined and impulsive-hyperactive subtypes predominate, unlike in industrialized countries where the impulsive-hyperactive type is the least common.[17,18]
On the other hand, in the Ten-Year Public Health Plan 2012-2021, as a priority dimension, within the objective "Zero tolerance with an avoidable disability", mental health with a component of prevention and comprehensive care for mental problems and disorders is

proposed. Within this component, the aim is to strengthen institutional and community management to guarantee comprehensive care for mental problems and disorders and associated events.[19] Given our social characteristics and prevalence of mental illnesses, it is essential to develop diagnostic support tools and evaluation of improvement for pathologies associated with poor impulse control about ADHD.[20] In this way, understanding the neurophysiological mechanisms of impulsivity is of vital importance for a better understanding and management of ADHD.[21]
Additionally, the development of early detection tools based on neurophysiological biomarkers would allow for more efficient therapeutic actions in the prevention of ADHD chronification, thus improving the quality of life of patients from the initial phases of their life course.[22] In this sense, several attempts have been made to characterize ADHD using elements that provide an objective diagnosis and whose extraction is low cost. Among the experimental methods to try to answer this question are electrophysiological studies, mainly those based on EEG.[23] Despite the challenges presented by this type of signal given its high noise components, it has been used successfully in diagnostic tasks.[15] However, the success of the diagnosis is highly dependent on the quality of the data, so there is a need to characterize and identify key components of the EEG signals to identify a certain pathology.[24]
Thus, the literature shows that the most recurrent analyzes to address the EEG are the time-frequencies since there is a significant relationship between the EEG signals spectrum and human behavior, cognitive status, or mental illnesses.[25,26] Therefore, spectral analysis is a useful tool to analyze EEG.[27] Based on this, the development of a methodology for characterizing ADHD from EEG recordings by analyzing multiple time-frequency characteristics of brain electrical activity constitutes a promising approach to address this challenge. Consequently, diagnostic support alternatives are generated from the generation of added value from the characteristics of the signal and research from both a clinical and technological point of view.

2.3 State of the art

In practice, many paradigms can be used to validate EEG recordings[28] and implement the response inhibition that is a cognitive control function and typical studied with a stop-signal task paradigm. Stop-signal response time (SSRT) analyses how rapidly an already-initiated response can be canceled for example.[29] This documented implement reward stop-signal task (RSST)was implemented for EEG recording to use response inhibition signals.
To extract relevant features of a sampled signal the state of art proposes the following tools; The FFT allows extract the frequency components of a sampled signal using a very efficient algorithm to perform the DFT signal.[30] This tool has the disadvantage that lost time information i.e., cannot associate a frequency component with time when it occurred. To get time-frequency information are needed techniques that operate on short segments

of the signal to localize the frequencies in time, where when increase the time resolution causes a decrease in the frequency resolution and vice versa.[31] To extract time-frequency information from a time signal is possible to apply FFT in a time-limited window assuming that the signal is stationary on the window interval, this is called STFT.[32] Thus, to window width keep constant, exists a fixed time-frequency resolution. Also, this tool is used to extract features from neural disorders like autism beside to FFT, getting a score of 82.4%.[33] On the other hand, DWT is widely using in biomedical engineering areas and provides more flexible windows that change time-frequency resolution to get information about a signal by an efficient algorithm, being able to obtain a 100% accuracy for EEG signal classification tasks.[34] Thus, DWT is a signal analysis tool that allows extracting relevant features on the time-frequency domain.[35] Different works show that wavelet transform provides a successful score rate, being the Daubechies sets useful to perform time-frequency decomposition, getting spectrum features, and demonstrating a clinical application.[36] Thus, DWT is useful tool for medicine. Authors like[37] show that statistical features derived from DWT can classifier EEG signals with 100% of accuracy to identification of epileptic seizure. Works like[38] propose features like energy, entropy, and standard deviation extracted from the EEG signals using DWT being the energy a very important feature to reach an accuracy of about 91.2%. Other papers[39] explored how the multiresolution decomposition allows decomposing a signal into frequency bands to analyze spectral components of interest and implement a filtering process. As shown[40] to analyze ADHD, Daubechies wavelets are useful to eliminate noise and extract important information of EEG signals using 3 levels of decompositions to get 3 sets of details coefficients and 1 of approximation coefficients. In this form was possible to get 94.74% of accuracy in the classification task using k-nearest neighbor classifiers and 90.04% with Decision Tree classifier.
To perform the classification task about EEG signals can be using random forest classifier,[41] SVM or BPNN obtaining a classification accuracy of 90%,[42] or Logistic Regressor classifier that can reach an average accuracy of 95.88%.[43] The latter has the advantage to provide an easy interpretation. Authors like[44,45] show how a binomial logistic regression classifier can be used to analyze ADHD and obtain an interpretable result about inhibition tasks. Another advantage to using logistic regression classifier is that allows to know the probability that a subject with ADHD will score positively in the classification task,[46] useful in clinical diagnostic.

2.4 Objectives

2.4.1 General objective

Develop a methodology to characterize EEG signals using multiresolution analysis tools and a logistic regression classifier to support the diagnosis of attention deficit hyperactivity disorder ADHD.

2.4.2 Specific objectives

- Implement a methodology for the EEG signals from the selection of decomposition levels of interest through the discrete wavelet transform.
- Develop a methodology for feature selection of the EEG signals from the discrete wavelet transform.
- Implement a methodology to tuning a logistic regression classifier to improve and validate the classification task easing the interpretability of the results from the weights obtained.

Chapter 3

Methods

3.1 DWT to feature extraction

A DWT system considers two types of real and real-valued functions to decompose a single-channel signal, namely, a scaling function and a wavelet function to extract information from the identification of coherent structure. The scaling function is an energy function that has a zero frequency component and unit norm, denoted by $\varphi(t) \in L^2$, where L^2 represents the space of square-integrable functions. The scaling functions can be translated by integer steps k to create a set of orthonormal basis denoting as following:

$$\{\varphi_k(t)\} = \{\varphi_k(t) = \varphi(t-k);\ k \in \mathbb{Z},\ \langle \varphi_{k1}(t), \varphi_{k2}(t) \rangle = \delta[k1 - k2]\} \tag{3.1}$$

Being δ the discrete Dirac delta function. This set can span a subspace $Span\{\varphi_k(t)\} \subset L^2$. To increase the size of the subspace spanned, each element can be scaled in a j factor keeping unit norm, giving the following set:

$$\{\varphi_{j,k}(t)\} = \{\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k);\ j,k \in \mathbb{Z}^2,\ \langle \varphi_{j,k1}(t), \varphi_{j,k2}(t) \rangle = \delta[k1 - k2]\} \tag{3.2}$$

Notice that if the j value increases, the set of scaling function translations on this scale become more contracted, giving more resolution in the space spanned, i.e. the size of this space $Span_j\{\varphi_{j,k}(t)\}$ grows, thus:

$$\lim_{j \to \infty} \{\varphi_{j,k}(t)\} = L^2 \tag{3.3}$$

$$\lim_{j \to -\infty} \{\varphi_{j,k}(t)\} = \{\mathbf{0}\} \tag{3.4}$$

Therefore, any function $f(t) \in L^2$ can be approximated at j_0 scale as a linear combination of elements of the set $\{\varphi_{j_0,k}(t)\}$. This approximation is the projection of the function $f(t)$ in the space $Span_{j_0}\{\varphi_{j_0,k}(t)\}$. The approximation function at j_0 scale is noted by $f_{j_0}(t)$ and is calculated by:

$$f_{j_0}(t) = \sum_{k=-\infty}^{\infty} a_{j_0,k} \cdot \varphi_{j_0,k}(t) \tag{3.5}$$

Where $a_{j_0,k}$ represent the approximation coefficients at j_0 scale and they depict the amplitude of the function $\varphi_{j_0,k}(t)$ contained in $f_{j_0}(t)$.These coefficients can be calculated by the inner product between $\varphi_{j_0,k}(t)$ and $f(t)$:

$$a_{j_0,k} = \langle f(t), \varphi_{j_0,k}(t) \rangle = \int_{-\infty}^{\infty} f(t) \cdot \varphi_{j_0,k}(t) dt \tag{3.6}$$

Notice that:

$$\lim_{j_0 \to \infty} f_{j_0}(t) = f(t) \tag{3.7}$$

Hence, the scaling function set approaches the function $f(t)$ but is needed another set of functions to describe better the important features or coherent structures of $f(t)$.
The wavelet function is an energy function that no has a zero frequency component and unit norm, denoted by $\psi(t) \in L^2$. As the scaling functions, space $Span_j\{\psi_{j,k}(t)\}$ can be spanned with the orthonormal basis functions defined as:

$$\begin{aligned} \{\psi_{j,k}(t)\} = &\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k);\ j,k \in \mathbb{Z}^2, \\ &\langle \psi_{j1,k1}(t), \psi_{j2,k2}(t) \rangle = \delta[j1 - j2] \cdot \delta[k1 - k2]\} \end{aligned} \tag{3.8}$$

Let be $f(t) \in L^2$. This function can be rewritten in terms of elements of $Span_j\{\psi_{j,k}(t)\}$:

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(t) \tag{3.9}$$

Where $d_{j,k}$ represent the detail coefficients, they depict the amplitude of the function $\psi_{j,k}(t)$ contained in $f(t)$ and represent better its features. These coefficients can be calculated by the inner product between $\psi_{j,k}(t)$ and $f(t)$:

$$d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle = \int_{-\infty}^{\infty} f(t) \cdot \psi_{j,k}(t) dt \tag{3.10}$$

Set a scale of interest j_0, the equation (3.9) take the following form:

$$\begin{aligned} f(t) &= \sum_{j=-\infty}^{j_0-1} \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(t) \\ &= f_{j_0} + \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(t) \end{aligned} \tag{3.11}$$

Or more compact:

$$f(t) = f_{j_0} + \sum_{j=j_0}^{\infty} g_j(t) \tag{3.12}$$

being:

$$g_j(t) = \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(t) \tag{3.13}$$

Equation (3.12) establishes that any function in L^2 can be decomposed in an approximation function at j_0 scale and details functions $g_j(t)$. Replacing $f(t)$ by f_{j_0+1} in this equation:

$$\begin{aligned} f_{j_0+1} =& f_{j_0} + \sum_{j=j_0}^{j_0} g_j(t) \\ =& f_{j_0} + g_{j_0}(t) \end{aligned} \tag{3.14}$$

The above equation shows that the projection of a function in an upper scale space $Span_{j_0+1}\{\varphi_{j_0+1,k}(t)\}$ can be represented with the sum of its projection in lower scale spaces $Span_{j_0}\{\varphi_{j_0,k}(t)\}$ and $Span_{j_0}\{\psi_{j_0,k}(t)\}$. In general:

$$Span_{j_0+1}\{\varphi_{j_0+1,k}(t)\} = Span_{j_0}\{\varphi_{j_0,k}(t)\} \oplus Span_{j_0}\{\psi_{j_0,k}(t)\} \tag{3.15}$$

Thus, the wavelet set $\{\psi_{j_0,k}(t)\}$ is the complement of the scaling set $\{\varphi_{j_0,k}(t)\}$ to reach $\{\varphi_{j_0+1,k}(t)\}$. Also, as the wavelet and scaling functions at chosen scale reside in the immediately upper scaling space (the space with more resolution), the following equations can be established:

$$\varphi_{j-1,k}(t) = \sum_{\tau} h_0(\tau)\varphi_{j,k}(t-\tau) \tag{3.16}$$

$$\psi_{j-1,k}(t) = \sum_{\tau} h_1(\tau)\varphi_{j,k}(t-\tau) \tag{3.17}$$

So, the representation coefficients at any level of decomposition or detail are calculated from the coefficients of the immediately upper level of decomposition:

$$a_{j-1,k} = \sum_{m} h_0(m-2k)a_{j,k} \tag{3.18}$$

$$d_{j-1,k} = \sum_{m} h_1(m-2k)a_{j,k} \tag{3.19}$$

Where $h_0(\tau)$ and $h_1(\tau)$ are low pass and high pass filters, respectively, as shown in Figure 3.1.

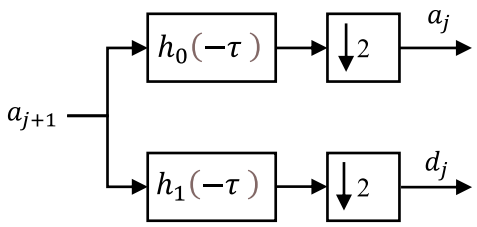


Figure 3.1. DWT decomposition of signal at different levels.

Thus, a single channel in an EEG trial $f_c(t) \in L^2$ can be written as a linear combination of the basis functions spanning $Span_j\{\varphi_{j,k}(t)\}$ and $Span_j\{\psi_{j,k}(t)\}$:

$$f_c(t) = \sum_{k=-\infty}^{\infty} a_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(t) \tag{3.20}$$

Where j_0 is the level of detail or decomposition of $f_c(t)$.
Due to the EEG signals were sampled, their scales and translations steps take finite values, being the samples the maximum resolution possible to reach, called the level of decomposition 0. Therefore, applying the DWT decomposition channel-wise in an EEG trial results in a set of coefficients representing the signal in the wavelet domain: $\boldsymbol{F}_i \rightarrow \{a_{ijk}, d_{ijk} : j \in [-J,0], k \in [0, \lceil T \cdot M_n \cdot 2^j \rceil]\}$, being J the number of considered decomposition levels, $i \in [1, M_n]$, and M_n the number of trials in the subject n.
Since M_n varies subject-wise, we extract a set of statistical features to build a fixed-length representing vector with physical interpretability. At each trial signal of the subject, is applicated DWT, return $J+1$ arrays correspond to J levels of detail coefficients and 1 of approximation coefficients.

3.2 Feature selection

At each decomposition level, including the last approximation, we compute five features: average coefficient, standard deviation, average peak location, the standard deviation of the peak location, and the maximum absolute coefficient. Let be c the coefficients array of interest (can be approximation coefficients or detail coefficients), the features can be determined as follows:

- **Average**

$$mean(c) = \frac{1}{N}\sum_{n=0}^{N-1} c[n] \tag{3.21}$$

- **Standard deviation**

$$std(c) = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}(c[n] - mean(c))^2} \tag{3.22}$$

To find the location of the peaks is needed finds all local maxima by a simple comparison of neighboring values as the following equation:

$$peak_{index}(c) = index\{c[n]\}\ if\ c[n-1] \leq c[n] \geq c[n+1]\ and\ c[n] > 0 \tag{3.23}$$

- **Average peak location**

$$mean(peak_{index}(c)) \tag{3.24}$$

- **Standard deviation of the peak location**

$$std(peak_{index}(c)) \tag{3.25}$$

- **Maximum absolute coefficient**

$$max_{abs}(c) = max\{|c|\} \tag{3.26}$$

As a result, is obtained a feature matrix $\boldsymbol{X} \in \mathbb{R}^{N \times P}$, being N the number of subjects and P=$5D(J+1)$ the total number of features.

3.3 Interpretable logistic regression classifier

The logistic regression classifier is a linear model of a binary classifier. This model uses a sigmoid function as an activation function. The sigmoid function denoted by $\sigma(z)$ is shown in Figure 3.2. This function $\sigma(z):\ \mathbb{R} \to (0,1)$ give the probability that a subject belongs to a specific class y, and is determined by the following equation:

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{3.27}$$

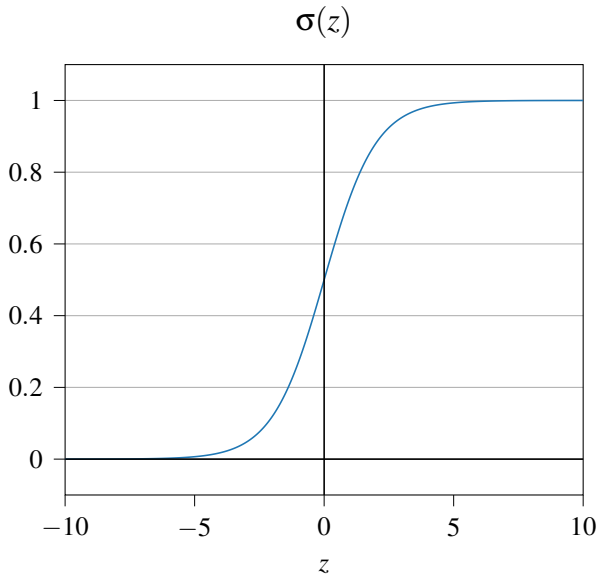


Figure 3.2. Sigmoid function

Let the matrix $\boldsymbol{X}$, where its rows $\boldsymbol{x}_n \in \mathbb{R}^P$ correspond to the features from the n-th sample, being labeled as $y_n \in \{0,1\}$ (0: Control and 1: ADHD). The logistic regression classifier classifies a sample vector according to the posterior probability for the target class $y = 1$ parameterized by a weighting vector $\boldsymbol{w} \in \mathbb{R}^P$ given by:

$$p(y = 1 \mid \boldsymbol{x}; \boldsymbol{w}) = \sigma\left(\boldsymbol{w}^\top \cdot \boldsymbol{x}\right) = \frac{1}{1 + e^{-\boldsymbol{w}^\top \cdot \boldsymbol{x}}} \tag{3.28}$$

In this case $z = \boldsymbol{w}^\top \cdot \boldsymbol{x}$ and p means the probability to $\boldsymbol{x}$ belong to the class $y = 1$ given the weighs $\boldsymbol{w}$. Figure 3.3 show how the logistic regression model gets an imput to give a class label.

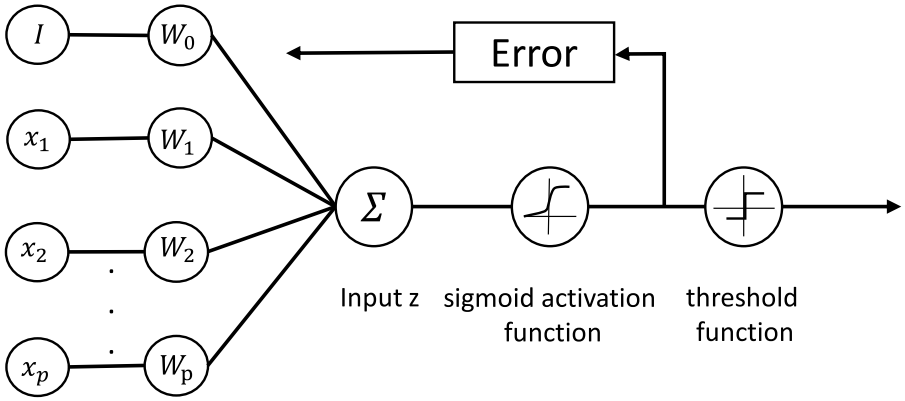


Figure 3.3. Logistic regression classifier model

To tune the weighting vector, the following optimization problem is stated that minimizes the cross-entropy loss constrained by the L_1 norm of $\boldsymbol{w}$:

$$\begin{aligned} \min_{\mathbf{w}} & -\sum_{n=1}^{N} y_n \log p\left(1 \mid \mathbf{x}_n; \boldsymbol{w}\right) + \left(1 - y_n\right) \log\left(1 - p\left(1 \mid \mathbf{x}_n; \boldsymbol{w}\right)\right) \\ & \text{subject to } \|\boldsymbol{w}\|_1 \leq C \end{aligned} \tag{3.29}$$

Being $C \in \mathbb{R}^+$ the inverse regularization parameter. The resulting weights select and rank the fewest number of features optimizing the classification task. Consequently, the selected wavelet-based features can be straightforwardly interpreted as the spatial locations and decomposition levels enhancing the ADHD discrimination.

Chapter 4

Experimets and results

4.1 Dataset and preprocessing

For evaluating the proposed methodology, was considered an EEG dataset obtained recorded during the execution of the *Reward Stop-Signal Task* (RSST). The RSST paradigm commands the participant to press a key when confronted with a frequent stimulus, labeled as *Go*, unless an infrequent stimulus appears after the *Stop* signal. The dataset holds 64 children diagnosed as either Healthy Control (HC) or ADHD. Each child executed four blocks of four minutes length, receiving a reward when succeeding in inhibiting the motor response after being presented with the Stop signal. The paradigm included a *Smiley* sticker reward, followed by a *Low* amount of candies and a *High* amount of candies. If the reward decreases from block to block the subject belongs to the decreasing condition (DC) group. On the contrary, when the reward increases from block to block the subject belongs to the increasing condition (IC) group. Only trials in which children pressed a key despite the Stop stimulus (failed inhibitions) constitutes the database to identify the pathological error monitoring associated with neuropsychiatric conditions.[47,48] Table 4.1 summarizes the number of children per diagnosis, along with the average number of failed inhibitions on each reward.

Table 4.1. Number of subjects and average number of failed inhibitions in the dataset.

Condition	Class	N	*Smiley*	*Low*	*High*
Decreasing	Control	14	09.0±02	8.45±04	7.03±04
	ADHD	15	12.9±04	13.0±05	14.0±06
Increasing	Control	17	10.0±07	9.05±05	8.01±05
	ADHD	17	14.6±06	14.0±04	13.3±05

Regarding the time series details, EEG signals were recorded at 250 Hz and 32 channels distributed over the scalp, registering activity over the medial frontal,[49] left frontal,[48] ventromedial orbitofrontal, and prefrontal cortices known to evoke event-related negative (ERN)

waves,[50] as show Figure 4.1. Each trial is trimmed 200 ms before and 800 ms after the *Go* stimulus, producing time series $\boldsymbol{F} \in \mathbb{R}^{D\times T}$ lasting T=250 time instants over D=32 channels.

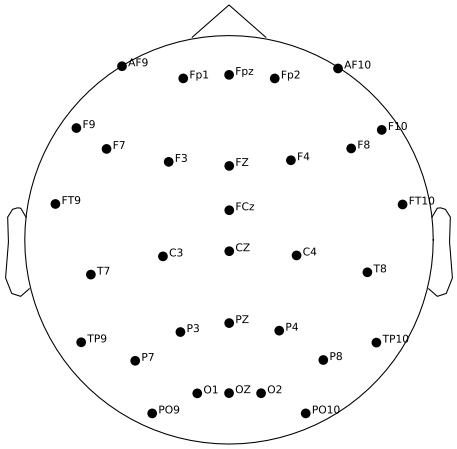


Figure 4.1. 10-20 EEG montage of RSST dataset.

4.2 DWT set selection

The proposed methodology uses a Daubechies 4 wavelet set in DWT as it has been widely used in the state of the art to highlight or detect relevant features in electrophysiological signals. Figure 4.2 shows a graph in time of scaling and wavelet functions of the set Daubechies 4 and their spectral density magnitude. Observe that the scaling function is a low pass filter and the wavelet function has a zero DC frequency component and is a pass band filter. Note that in the frequency domain there is spectral leakage because the signals in time present a compact support equal to 3.

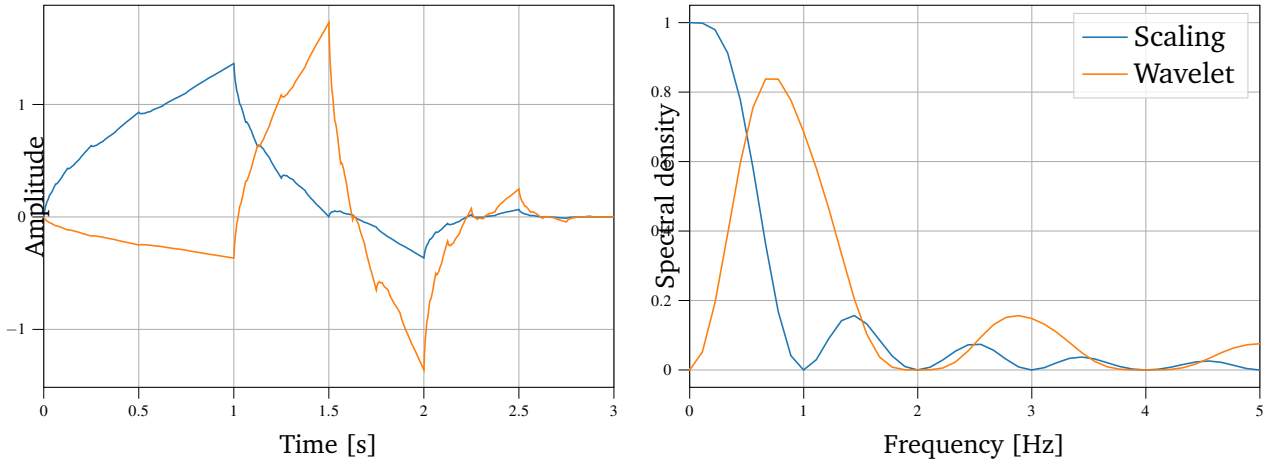


Figure 4.2. Daubechies 4 set in time and frequency

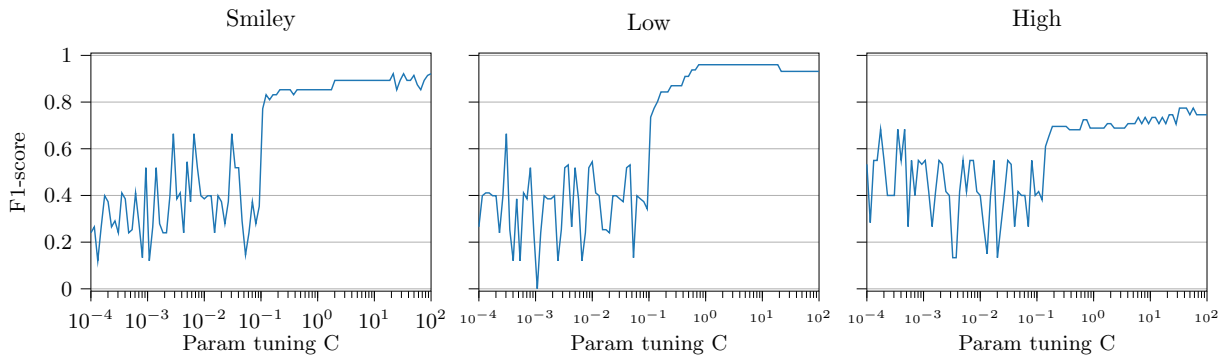
To identify the best signal decomposition level to highlight discriminative features between ADHD subjects and controls were compared the classification results of the first five decomposition levels with coefficients of details and the last level of approximation. These features feed the logistic regression classifier using statistical moments and spike descriptors in a 5-fold cross-validated grid search within $C \in [10^{-4}, 10^{2}]$. Table 4.2 shows the average f1-score achieved over the test set along with C for each of the reward and conditions, highlighting that for the proposed approach, the best performances were obtained with the level 3 decomposition in *IC* condition, exhibiting the highest scores and lowest standard deviations for the three rewards than *DC* condition. Contrary, in the DC condition, the best score is not in the same level of decomposition for the three rewards. Although in both conditions the best performance is *Low* reward.

Table 4.2. Classification results of the F1 score average for each decomposition level and reward.

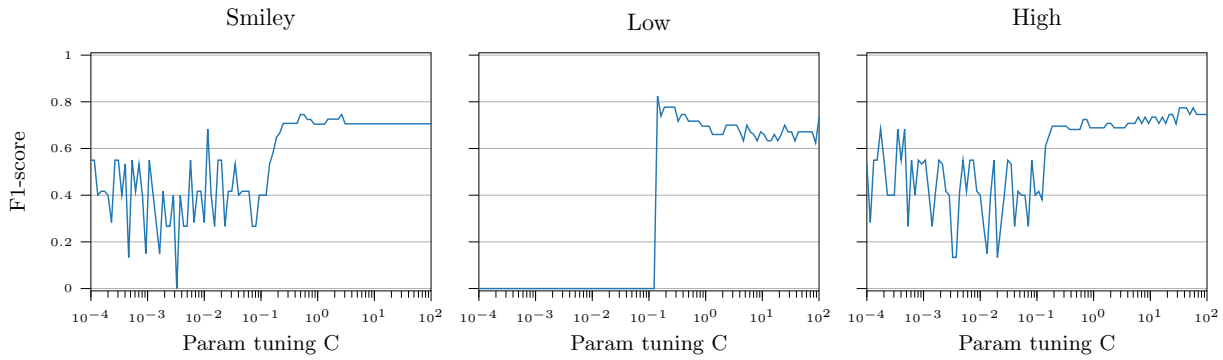
Conditions	Rewards	Decomposition levels				
		1	2	3	4	5
IC	*Smiley*	88±10	86±12	**94±07**	91±11	85±08
	Low	91±11	91±11	**96±08**	96±08	93±09
	High	86±08	85±08	**88±07**	85±08	88±10
DC	*Smiley*	71±14	73±21	71±18	**75±18**	72±15
	Low	**82±11**	76±22	78±14	79±03	78±03
	High	75±07	**77±21**	74±16	75±18	76±17

4.3 Parameter tuning

Starting from the best level of decomposition shown in Table 4.2 and to find the set of parameters that will improve the classification. We fitted the inverse of the regularization parameter C in a grid search as shown in Figure 4.3. It is worth noting that, for values of $C > 10^{-1}$, the f1-score increases for all rewards and conditions because are needed a certain amount of non-zero weights to reach the best value. Also, the best score is always obtained in the *IC* condition as was mentioned before. Therefore, the best performances are obtained with low regularization values (the inverse value of C).Note that *Low* reward obtained the best performance in two conditions.



(a) IC condition

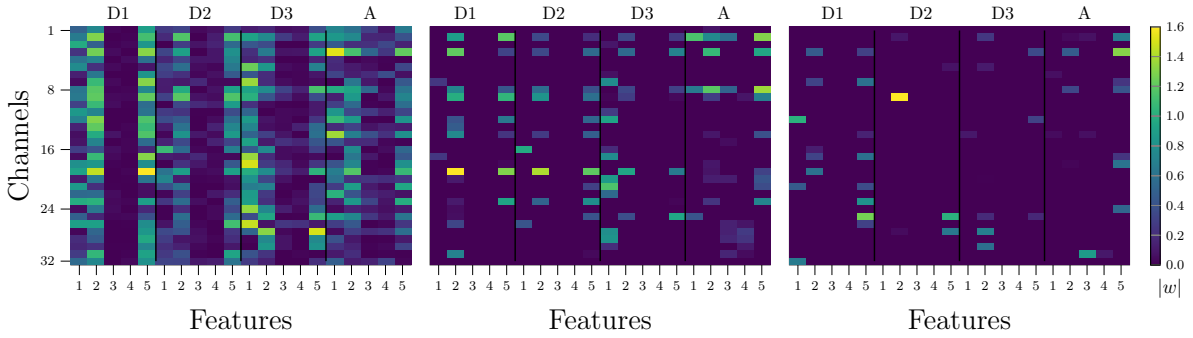


(b) DC condition

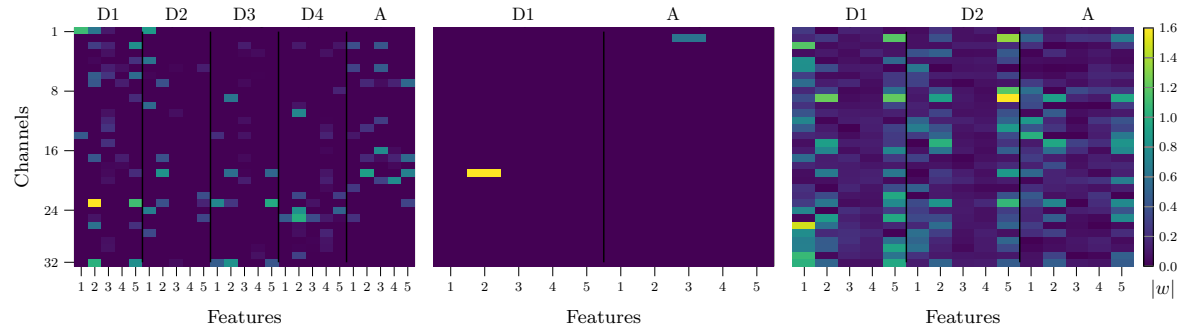
Figure 4.3. *C* tuning curve. Five test folds are averaged for each reward and condition

4.4 Feature extraction and performance

Figure 4.4 shows the magnitude of the weights of channel features of each condition and reward associated with the level of decomposition that reaches the best f1-score value. Notice that features 3 and 4 (mean of peak indexes and standard deviation of peak indexes respectively) have near-zero values which suggest that few relevant to the classification task in all cases. Beside, IC smiley and DC high present the most numbers of weighs different to zero, which is associated with a high value of *C* or low value of regularization. This means that are needed many features to reach the best score. On the other hand, DC low has a strong regularization, for this reason, it presents only two non-zero weights and the magnitude of its weight in channel 20 (**FT9**) has a high magnitude. In this case, only are needed two feature to get the best classification score.



(a) IC condition



(b) DC condition

Figure 4.4. Magnitude of the weights associated with the level of decomposition that reaches the best f1-score value

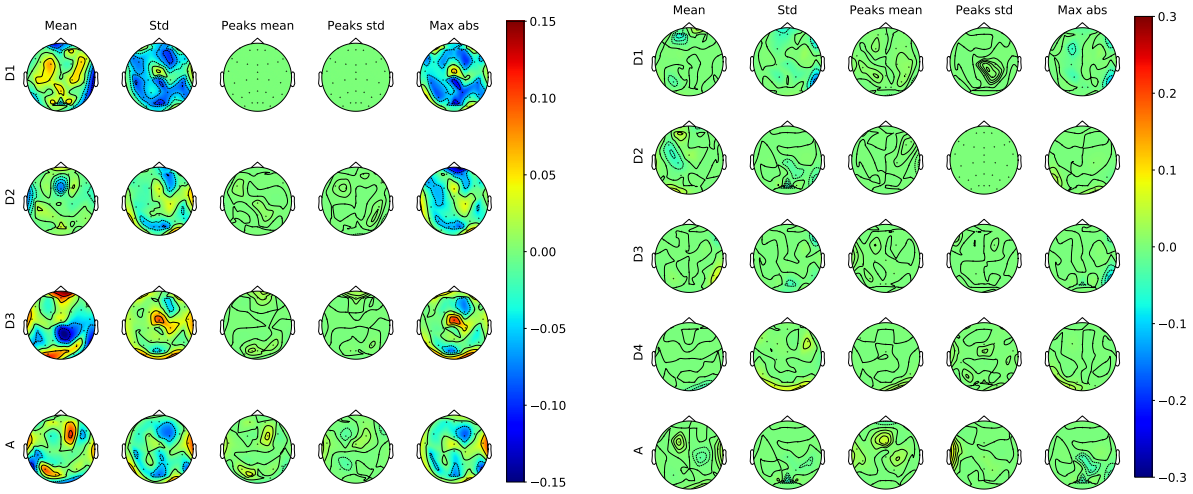
Table 4.3 summarizes the average f1-score, precision, recall, and AUC with standard deviation obtained with the logistic regression classifier best parameters for the corresponding decomposition levels for all conditions and reward. Results show that the best performance is almost on *IC* condition with the best performance on *Low* reward with the shortest deviations obtaining a more reliable classification performance. This latter is characterized by not presenting false positives, that is, control patients such as ADHD. On DC condition the best scores are on Low and High reward being Smiley reward ever the worst and present a very low AUC performance. The other rewards have high-performance measures on precision and AUC, which means that they present few false positives being High reward the best, but Low reward improves F1 and recall on this condition.

Table 4.3. Performance measures for all rewards and conditions in best decomposition level and *C* parameter.

Conditions	Rewards	Performance measures			
		F1	*Precision*	*Recall*	*AUC*
IC	*Smiley*	94±07	95±10	95±10	98±03
	Low	**96±08**	**100±0**	**93±13**	**100±0**
	High	88±07	86±12	93±13	93±06
DC	*Smiley*	75±18	85±20	73±25	66±27
	Low	**82±11**	88±15	**80±16**	89±17
	High	77±21	**90±20**	73±25	**91±13**

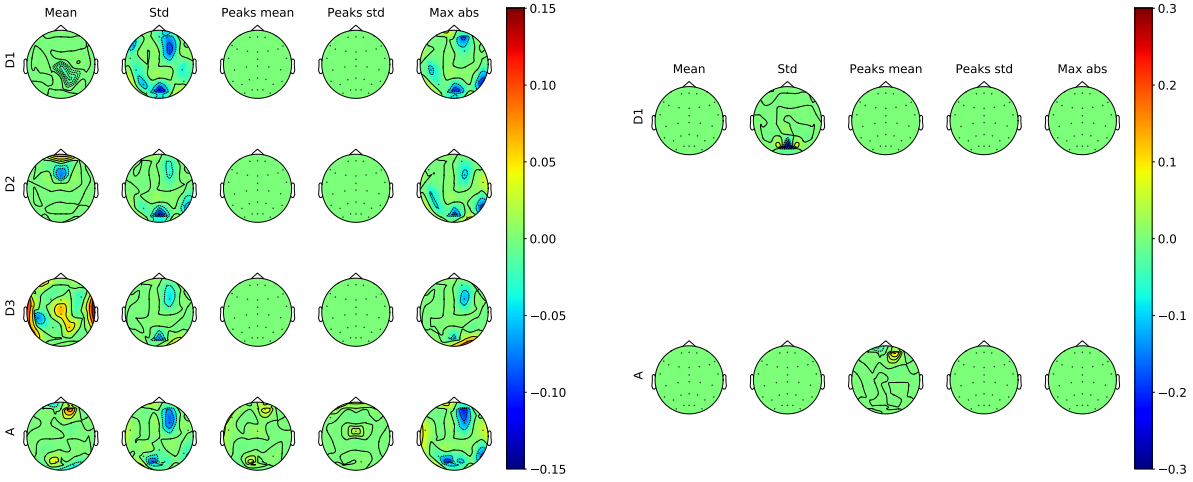
4.5 Interpretation analysis

Figures 4.5 to 4.7 shows the weights associated with each feature in correspondence with the channel and coefficient from which they were extracted for all conditions and rewards over the head scalp according to their best decomposition level. Red and blue regions indicate leading contributions to the discrimination, either by a positive or negative weight, respectively. In turn, green areas denote close-to-zero weights and a lack of effect on the classifier prediction. Note that the peak features are less relevant for the classification task. In DC condition the *High* reward presents the most relevant weight, being zero or near to zero in the other cases except for a few small areas. Also, in this reward exist a high activity in the frontal lobule on detail coefficients on level 1 of decomposition and mean feature. These coefficients are associated with high frequencies (more than 62.5 Hz) and are needed for the classification task. On the other hand, on IC condition is shown fewer intensities weighs being High reward again wich presents the most insensitive values.



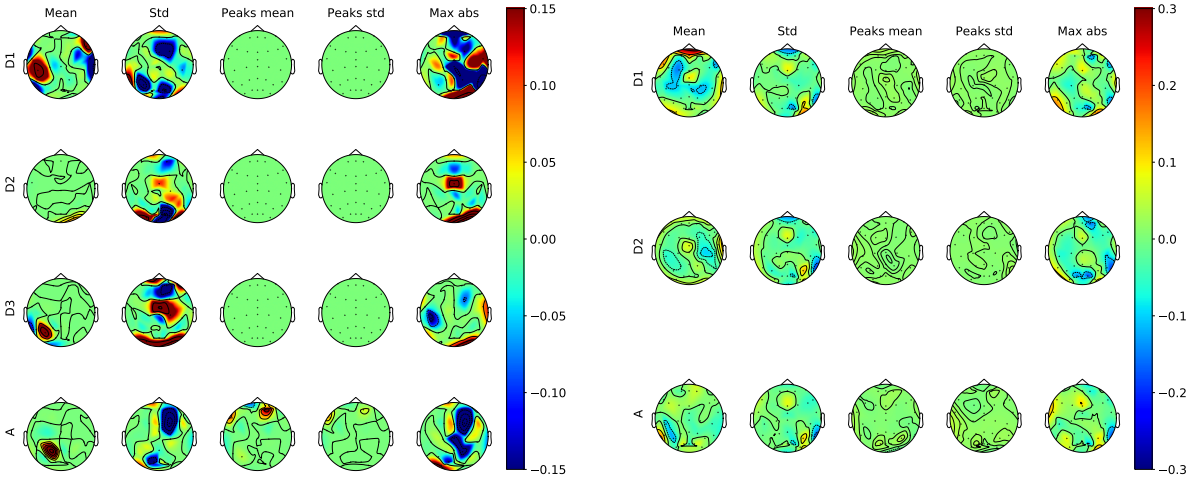
(a) IC (b) DC

Figure 4.5. Spatial interpretation of the weights obtained from the logistic regression classifier for each feature and best decomposition level at all reward and condition for smiley reward



(a) IC (b) DC

Figure 4.6. Spatial interpretation of the weights obtained from the logistic regression classifier for each feature and best decomposition level at all reward and condition for low reward



(a) IC (b) DC

Figure 4.7. Spatial interpretation of the weights obtained from the logistic regression classifier for each feature and best decomposition level at all reward and condition for high reward

Chapter 5

Conclusions

This work proposes an ADHD-supported diagnosis methodology through EEG recording decomposition using the Daubechies 4 wavelet set to perform DWT. According to the RSST paradigm, the proposed methodology extracts characteristics of mean, standard deviation, and maximum magnitude of the signal, and mean and standard deviation of the index of peaks for each decomposition level of signal and type of condition and reward, during failed inhibition trials. Then, these characteristics feed a logistic regression classifier with l1 regularization facilitating the interpretability of the classification results needed in this type of clinical task.

To identify the optimal decomposition for each of the three rewards, we inspected up to five levels. Each tested number of decomposition levels included all detail coefficients and the last ones for approximation. Performance metrics in Table 4.2 show that the best number of decomposition levels is $J = 3$. Considering that the frequencies captured at the third decomposition level range from 15.6Hz to 31.2Hz, the methodology highlights the role of the Beta brain rhythm as an ADHD biomarker, agreeing with the clinical literature.[51,52] Besides, including higher decomposition levels introduces noisy features, so hampering the classification performance.

In the parameter tuning, we searched the optimal box constraint C for each reward. From Figure 4.3, it is worth noting that the F1 score considerably increases when $C > 0.1$ for the three reward levels, implying a low regularization requirement for the classification problem. Nonetheless, the *High* reward, which underperforms *Smiley* and *Low*, results in the most constrained feature set (the shortest C value), indicating the lowest discrimination capability of such a reward. In the case of Smiley, the best scores emerge at very high C values, so that regularizing hampers the supported diagnosis. Consequently, the *Smiley* reward can not highlight discriminative features in scale or space. In turn, *Low* reaches the highest scores with a balanced C, becoming the reward level evoking the widest gap between both diagnoses while selecting a subset of relevant interpretable features for the task.

Thanks to information provided by the logistic regressor and to obtain information on

the characteristics used, we visualized the weight of the five features at the best level of decomposition in two conditions, the Figure 4.4 shows that the features of mean and deviation of the peaks do not contribute to the classification. This suggests that discrimination between classes is not associated with a temporal factor, but is related to the energy distribution of the signal being the most relevant statistical descriptors in the classification task in both conditions. Besides, from the results shown in Figures 4.5 to 4.7 we identified that Low reward in IC only highlights specific areas that are relevant to classification being the most important the associated with D3 coefficients, so the coefficients associated with a signal representation without high-frequency components improve feature discrimination because, in level 3 of decomposition, a bandwidth of 15.625 to $31.25Hz$ of frequency information of the signal is found, which is related to a Beta band of EEG signals. So the methodology highlights the role of the Beta brain rhythm as an ADHD biomarker, agreeing with the clinical literature.[51,52] Besides, including higher decomposition levels introduces noisy features, so hampering the classification performance.

From the RSST paradigm, our findings support the hypothesis that the motivational effect influences the inhibitory control of children with ADHD, as reflected in Table 4.3 for both conditions. ADHD children engage differently than controls in the inhibitory task when receiving a *Smiley* or a *Low* reward. The decay on the performance under *High* suggests that powerful rewards either smoothen the anomalous error monitoring response of ADHD children or saturates the motivational effect for both HC and ADHD.

For future work, we devise three research directions: Firstly, we will extend the methodology to other feature sets, e.g. connectivity measures, to gain knowledge about the disorder. Secondly, we will analyze other cognitive tasks targeting clinical ADHD symptoms on the space, time, and frequency domains, aiming at optimizing the paradigm, EEG recording montage, and feature set for supported diagnosis of ADHD. Thirdly, we will introduce a subject-wise interpretable machine learning strategy, such as the Shapley additive explanations, so that the model not only suggests a diagnosis but also describes the EEG features and values leading it.

Bibliography

[1] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[2] S. M. Snyder, T. A. Rugino, M. Hornig, and M. A. Stein, "Integration of an EEG biomarker with a clinician's ADHD evaluation," *Brain and Behavior*, 2015.

[3] M. Ahmadlou and H. Adeli, "Wavelet-synchronization methodology: A new approach for EEG-based diagnosis of ADHD," *Clinical EEG and Neuroscience*, 2010.

[4] J. N. Swartwood, M. O. Swartwood, J. F. Lubar, and D. L. Timmermann, "Eeg differences in adhd-combined type during baseline and cognitive tasks," *Pediatric Neurology*, vol. 28, no. 3, pp. 199–204, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0887899402005143

[5] D. F. Hermens, M. R. Kohn, S. D. Clarke, E. Gordon, and L. M. Williams, "Sex differences in adolescent adhd: findings from concurrent eeg and eda," *Clinical neurophysiology*, vol. 116, no. 6, pp. 1455–1463, 2005.

[6] J. F. Saad, M. R. Kohn, S. Clarke, J. Lagopoulos, and D. F. Hermens, "Is the Theta/Beta EEG Marker for ADHD Inherently Flawed?" *Journal of Attention Disorders*, vol. 22, no. 9, pp. 815–826, 2018.

[7] M. J. Groom, G. Scerif, P. F. Liddle, M. J. Batty, E. B. Liddle, K. L. Roberts, J. D. Cahill, M. Liotti, and C. Hollis, "Effects of motivation and medication on electrophysiological markers of response inhibition in children with attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 67, no. 7, pp. 624–631, 2010.

[8] S. Whitmont, R. Meares, E. Gordon, I. Lazzaro, and S. Clarke, "The Modulation of Late Component Event Related Potentials by Pre-Stimulus EEG Theta Activity in ADHD," *International Journal of Neuroscience*, vol. 107, no. 3-4, pp. 247–264, 2008.

[9] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification of eeg signals using the wavelet transform," *Signal processing*, vol. 59, no. 1, pp. 61–72, 1997.

[10] R. Gabriel, M. M. Spindola, A. Mesquita, and A. Z. Neto, "Identification of adhd cognitive pattern disturbances using eeg and wavelets analysis," in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2017, pp. 157–162.

[11] H. A. Darwish, M. Hesham, A.-M. I. Taalab, and N. M. Mansour, "Close accord on dwt performance and real-time implementation for protection applications," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2174–2183, 2010.

[12] P. A. M. Kanda, "Análise de wavelets com máquina de vetor de suporte no eletrencefalograma da doença de alzheimer," Ph.D. dissertation, Universidade de São Paulo, 2012.

[13] R. C. Joy, S. T. George, A. A. Rajan, and M. Subathra, "Detection of attention deficit hyperactivity disorder from eeg signal using discrete wavelet transform," in *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*. IEEE, 2019, pp. 1–5.

[14] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 44, pp. 22 071–22 080, 2019.

[15] F. H. Duffy, A. Shankardass, G. B. McAnulty, and H. Als, "A unique pattern of cortical connectivity characterizes patients with attention deficit disorders: a large electroencephalographic coherence study," *BMC medicine*, vol. 15, no. 1, p. 51, 2017.

[16] C. Montiel-Nava, J. A. Peña, M. López, M. Salas, J. R. Zurga, I. Montiel-Barbero, D. Pirela, and J. J. Cardozo, "Estimaciones de la prevalencia del trastorno por déficit de atención-hiperactividad en niños Marabinos," *Revista de Neurologia*, 2002.

[17] C. Gómez-Restrepo, C. de Santacruz, M. N. Rodriguez, V. Rodriguez, N. Tamayo Martínez, D. Matallana, and L. M. Gonzalez, "Encuesta Nacional de Salud Mental Colombia 2015. Protocolo del estudio," *Revista Colombiana de Psiquiatria*, vol. 45, pp. 2–8, 2016.

[18] C. Hidalgo-López, A. M. Gómez-Álzate, J. García-Valencia, and J. D. Palacio-Ortiz, "Risk of attention deficit/hyperactivity disorder (ADHD) and other psychiatric disorders in siblings of ADHD probands," *Revista Colombiana de Psiquiatría (English ed.)*, vol. 48, no. 1, pp. 44–49, 2019. [Online]. Available: http://dx.doi.org/10.1016/j.rcpeng.2017.06.006

[19] Minisalud, "Dimension de vida saludable," p. 82, 2013. [Online]. Available: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/IMP_4feb+ABCminsalud.pdf

[20] N. Tamayo Martínez, C. J. Rincón Rodríguez, C. de Santacruz, N. Bautista Bautista, J. Collazos, and C. Gómez–Restrepo, "Problemas mentales, trastornos del afecto y de ansiedad en la población desplazada por la violencia en Colombia, resultados de la Encuesta Nacional de Salud Mental 2015," *Revista Colombiana de Psiquiatría*, vol. 45, pp. 113–118, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034745016300889

[21] J. Berg, R. Latzman, N. Bliwise, and S. Lilienfeld, "Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the upps for psychopathology," *Psychological assessment*, vol. 27, 03 2015.

[22] F. di Michele, L. Prichep, E. John, and R. Chabot, "The neurophysiology of attention-deficit/hyperactivity disorder," *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 58, pp. 81–93, 11 2005.

[23] A. Lenartowicz and S. K. Loo, "Use of EEG to diagnose ADHD." *Current psychiatry reports*, vol. 16, no. 11, p. 498, nov 2014.

[24] Z. Ghahramani, "Ghahramani 2015 Nature," *Nature*, vol. 27, pp. 452–459, 2015. [Online]. Available: https://www.repository.cam.ac.uk/bitstream/handle/1810/248538/Ghahramani2015Nature.pdf

[25] O. Dressler, G. Schneider, G. Stockmanns, and E. F. Kochs, "Awareness and the EEG power spectrum: analysis of frequencies," *British Journal of Anaesthesia*, vol. 93, no. 6, pp. 806–809, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000709121735794X

[26] B. J. Martínez-Briones, T. Fernández-Harmony, N. Garófalo Gómez, R. J. Biscay-Lirio, and J. Bosch-Bayard, "Working Memory in Children with Learning Disorders: An EEG Power Spectrum Analysis." *Brain sciences*, vol. 10, no. 11, nov 2020.

[27] S. Amin, R. Azim, T. Latif, A. Hoque, and F. M. Hasan, "Spectral analysis of human sleep EEG signal," *ICSPS 2010 - Proceedings of the 2010 2nd International Conference on Signal Processing Systems*, vol. 3, pp. 106–110, 2010.

[28] D. Bacher, A. Amini, D. Friedman, W. Doyle, S. Pacia, and R. Kuzniecky, "Validation of an EEG seizure detection paradigm optimized for clinical use in a chronically implanted subcutaneous device," *Journal of Neuroscience Methods*, vol. 358, no. June 2020, p. 109220, 2021. [Online]. Available: https://doi.org/10.1016/j.jneumeth.2021.109220

[29] C. N. Boehler, H. Schevernels, J. M. Hopf, C. M. Stoppel, and R. M. Krebs, "Reward prospect rapidly speeds up response inhibition via reactive control," *Cognitive, Affective and Behavioral Neuroscience*, vol. 14, no. 2, pp. 593–609, 2014.

[30] Maan M. Shaker, "EEG Waves Classifier using Wavelet Transform and Fourier Transform Maan M. Shaker International," vol. 1, no. 3, pp. 169–174, 2007.

[31] K. Samiee, P. Kovács, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 541–552, 2015.

[32] D. Yu, W. Jinzhen, S. Shaoying, and C. Zengping, "Detection of LFM signals in low SNR based on STFT and wavelet denoising," *ICALIP 2014 - 2014 International Conference on Audio, Language and Image Processing, Proceedings*, pp. 921–925, 2015.

[33] H. Behnam, A. Sheikhani, M. R. Mohammadi, M. Noroozian, and P. Golabi, "Analyses of EEG background activity in Autism disorders with fast Fourier transform and short time Fourier measure," *2007 International Conference on Intelligent and Advanced Systems, ICIAS 2007*, pp. 1240–1244, 2007.

[34] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," *Signal, Image and Video Processing*, vol. 8, no. 7, pp. 1323–1334, 2014.

[35] S. H. Lee, B. Abibullaev, W. S. Kang, Y. Shin, and J. An, "Analysis of attention deficit hyperactivity disorder in EEG using wavelet transform and self organizing maps," *ICCAS 2010 - International Conference on Control, Automation and Systems*, pp. 2439–2442, 2010.

[36] B. De Celis Alonso, J. M. Hernández López, J. G. Suárez García, and E. Moreno Barbosa, "A minireview on the use of wavelet analyses on physiological signals to diagnose and characterize ADHD," *International Journal of Basic and Applied Sciences*, vol. 6, no. 3, p. 57, 2017.

[37] T. Capri, E. Santoddi, and R. A. Fabio, "Multi-Source Interference Task paradigm to enhance automatic and controlled processes in ADHD," *Research in Developmental Disabilities*, vol. 97, no. August 2018, p. 103542, 2020. [Online]. Available: https://doi.org/10.1016/j.ridd.2019.103542

[38] R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengthe, P. R. Pal, and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction," *International Conference on Systems in Medicine and Biology, ICSMB 2010 - Proceedings*, no. December, pp. 405–408, 2010.

[39] H. Peng, B. Hu, Q. Shi, M. Ratcliffe, Q. Zhao, Y. Qi, and G. Gao, "Removal of ocular artifacts in EEG - An improved approach combining DWT and ANC for portable applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 600–607, 2013.

[40] H. T. Tor, C. P. Ooi, N. S. Lim-Ashworth, J. K. E. Wei, V. Jahmunah, S. L. Oh, U. R. Acharya, and D. S. S. Fung, "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105941, 2021. [Online]. Available: https://doi.org/10.1016/j.cmpb.2021.105941

[41] E. Saifutdinova, D. U. Dudysová, L. Lhotská, V. Gerla, and M. MacAš, "Artifact Detection in Multichannel Sleep EEG using Random Forest Classifier," *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, no. Cv, pp. 2803–2805, 2019.

[42] T. K. Shri, N. Sriraam, and V. Bhat, "Characterization of EEG signals for identification of alcoholics using ANOVA ranked approximate entropy and classifiers," *Proceedings of International Conference on Circuits, Communication, Control and Computing, I4C 2014*, no. November, pp. 109–112, 2014.

[43] H. Rajaguru, "Classification of Epilepsy from EEG Signals," pp. 577–580, 2017.

[44] A. Vélez-van Meerbeke, I. P. Zamora, G. Guzmán, B. Figueroa, C. A. López Cabra, and C. Talero-Gutiérrez, "Evaluación de la función ejecutiva en una población escolar con síntomas de déficit de atención e hiperactividad," *Neurologia*, vol. 28, no. 6, pp. 348–355, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.nrl.2012.06.011

[45] C. Rivas-Juesas, J. G. de Dios, M. Benac-Prefaci, and J. Colomer-Revuelta, "Análisis de los factores ligados al diagnóstico del trastorno por déficit de atención e hiperactividad en la infancia," *Neurologia*, vol. 32, no. 7, pp. 431–439, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.nrl.2016.01.006

[46] V. Richarte, M. Corrales, M. Pozuelo, J. Serra-Pla, P. Ibáñez, E. Calvo, M. Corominas, R. Bosch, M. Casas, and J. A. Ramos-Quiroga, "Spanish validation of the adult Attention Deficit/Hyperactivity Disorder Rating Scale (ADHD-RS): Relevance of clinical subtypes," *Revista de Psiquiatría y Salud Mental (English Edition)*, vol. 10, no. 4, pp. 185–191, 2017.

[47] Y. Groen, A. A. Wijers, L. J. Mulder, B. Waggeveld, R. B. Minderaa, and M. Althaus, "Error and feedback processing in children with adhd and children with autistic spectrum disorder: an eeg event-related potential study," *Clinical neurophysiology*, vol. 119, no. 11, pp. 2476–2493, 2008.

[48] C. S. van Meel, D. J. Heslenfeld, J. Oosterlaan, and J. A. Sergeant, "Adaptive control deficits in attention-deficit/hyperactivity disorder (adhd): the role of error processing," *Psychiatry research*, vol. 151, no. 3, pp. 211–220, 2007.

[49] K. R. Ridderinkhof, M. Ullsperger, E. A. Crone, and S. Nieuwenhuis, "The role of the medial frontal cortex in cognitive control," *science*, vol. 306, no. 5695, pp. 443–447, 2004.

[50] H. Garavan, T. Ross, K. Murphy, R. Roche, and E. Stein, "Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction," *Neuroimage*, vol. 17, no. 4, pp. 1820–1829, 2002.

[51] D. McAuliffe, K. Hirabayashi, J. H. Adamek, Y. Luo, D. Crocetti, A. S. Pillai, Y. Zhao, N. E. Crone, S. H. Mostofsky, and J. B. Ewen, "Increased mirror overflow movements in adhd are associated with altered eeg alpha/beta band desynchronization," *European Journal of Neuroscience*, vol. 51, no. 8, pp. 1815–1826, 2020.

[52] J. F. Saad, M. R. Kohn, S. Clarke, J. Lagopoulos, and D. F. Hermens, "Is the theta/beta eeg marker for adhd inherently flawed?" *Journal of attention disorders*, vol. 22, no. 9, pp. 815–826, 2018.